

PATENT

\* ATTORNEY DOCKET NO.: 040894-5726

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

Kiyoshi YAJIMA, et al.

Application No.: 09/964,677

Filed: September 28, 2001

)  
)  
)  
)  
)  
)

Group Art Unit: 2644

Examiner: Unassigned

For: VOICE RECOGNITION SYSTEM

Commissioner for Patents  
Washington, D.C. 20231

**CLAIM FOR PRIORITY**

Under the provisions of 35 U.S.C. §119, Applicants hereby claim the benefit of the filing date of Certified copy of Japanese Patent Application No. 2000-298536 filed September 29, 2000 for the above-identified United States Patent Application.

In support of Applicants' claim for priority, filed herewith is a certified copy of the Japanese application.

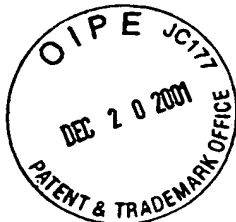
Respectfully submitted,

**MORGAN, LEWIS & BOCKIUS LLP**

Robert J. Goodell  
Reg. No. 41,040

Dated: December 20, 2001

**MORGAN, LEWIS & BOCKIUS LLP**  
1800 M Street, N.W.  
Washington, D.C. 20036  
(202)467-7000



日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2 0 0 0 年 9 月 2 9 日

出 願 番 号

Application Number:

特 願 2 0 0 0 - 2 9 8 5 3 6

出 願 人

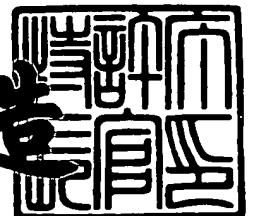
Applicant(s):

パイオニア株式会社

2 0 0 1 年 8 月 3 日

特 許 庁 長 官  
Commissioner,  
Japan Patent Office

及 川 耕 造



出 証 番 号 出 証 特 2 0 0 1 - 3 0 6 9 0 4 9

【書類名】 特許願

【整理番号】 54P0699

【提出日】 平成12年 9月29日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 3/00  
G10L 5/00  
G10L 7/00

【発明者】

【住所又は居所】 埼玉県鶴ヶ島市富士見6丁目1番1号 パイオニア株式会社総合研究所内

【氏名】 谷島 潔

【発明者】

【住所又は居所】 埼玉県鶴ヶ島市富士見6丁目1番1号 パイオニア株式会社総合研究所内

【氏名】 外山 聡一

【特許出願人】

【識別番号】 000005016

【氏名又は名称】 パイオニア株式会社

【代理人】

【識別番号】 100063565

【弁理士】

【氏名又は名称】 小橋 信淳

【手数料の表示】

【予納台帳番号】 011659

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

特 2 0 0 0 - 2 9 8 5 3 6

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声認識システム

【特許請求の範囲】

【請求項 1】 音声情報に基づいて生成された標準ベクトルを有する標準音響モデルと、

指定テキストに対応する発話となされると、前記発話された音声からノイズを除去した第 1 の特徴ベクトル系列を生成する第 1 の特徴ベクトル生成手段と、

前記発話された声からノイズを除去することなく第 2 の特徴ベクトル系列を生成する第 2 の特徴ベクトル生成手段と、

前記第 1 及び第 2 の特徴ベクトル生成手段で生成される前記第 1, 第 2 の特徴ベクトル系列に基づいて、前記標準ベクトルを補償した適応ベクトルを生成することで、前記指定テキストに対応して発話された声に適応した話者適応音響モデルを作成する作成手段と、

を具備することを特徴とする音声認識システム。

【請求項 2】 前記作成手段は、前記第 1 の特徴ベクトル生成手段で生成される前記第 1 の特徴ベクトル系列と前記標準ベクトルとの対応関係を調べることでパス探索を行うと共に、

前記第 2 の特徴ベクトル生成手段で生成される前記第 2 の特徴ベクトル系列と前記標準ベクトルとの対応付けを前記パス探索の結果に応じて行うことにより、前記適応ベクトルを生成することを特徴とする請求項 1 記載の音声認識システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声認識システムに関し、特に、ノイズに対してロバストな話者適応型の音声認識システムに関する。

【0002】

【従来の技術】

従来、話者適応型の音声認識システムとして、例えば図 9 に示す構成のものが

知られている。

【0003】

この音声認識システムは、予め作成された不特定話者の標準音響モデル100を備え、特定話者の発話した入力音声Scの特徴ベクトルと標準音響モデル100とを用いて話者適応音響モデル200を作成し、その特定話者の声に適応させて音声認識を行うようになっている。

【0004】

適応音響モデル200を作成する際、指定テキスト（文章や音節等）Txに該当する標準ベクトルVaを、標準音響モデル100からパス探索部4と話者適応部5に供給すると共に、実際に指定テキストTxを特定話者に発話させてその入力音声Scを入力する。

【0005】

そして、加法性ノイズ除去部1が入力音声Sc中に含まれる加法性ノイズを除去した後、特徴ベクトル生成部2が入力音声Scの特徴量である特徴ベクトル系列Vcfを生成し、更に乗法性ノイズ除去部3が特徴ベクトル系列Vcfから乗法性ノイズを除去することで、加法性ノイズと乗法性ノイズを除去した特徴ベクトル系列Vcを生成して、パス探索部4と話者適応部5に供給する。

【0006】

こうして、標準ベクトルVaと実際に発話された入力音声Scの特徴ベクトル系列Vcがパス探索部4と話者適応部5に供給されると、パス探索部4が、特徴ベクトル系列Vcを標準ベクトルVaと対比して、特徴ベクトル系列Vcの音節毎の出現確率と、音節から音節への状態遷移確率とを求めた後、話者適応部5が、上記の出現確率と状態遷移確率とに基づいて標準ベクトルVaを補償することで、特定話者固有の声の特徴に適応した話者適応音響モデル200を作成する。

【0007】

そして、話者適応音響モデル200を特定話者の声に適応させた後、特定話者が任意の発話をする、その発話音声の特徴ベクトルと話者適応音響モデル200の適応ベクトルとを照合し、最大尤度を与える話者適応音響モデル200を認識結果とすることで音声認識を行うようになっている。

【0008】

【発明が解決しようとする課題】

ところで、上記従来の適応型音声認識システムでは、適応音響モデル200を作成する際、加法性ノイズ除去部1は、スペクトルサブトラクション法によって加法性ノイズを除去し、乗法性ノイズ除去部3は、CMN法 (cepstrum mean normalization) によって乗法性ノイズを除去することで、ノイズの影響を受けない話者適応音響モデル200を作成することとしている。

【0009】

すなわち、加法性ノイズ除去部1は、入力音声Scのスペクトルを求めた後、そのスペクトルから加法性ノイズのスペクトルを除去し、乗法性ノイズ除去部3は、入力音声Scのケプストラム (cepstrum) の時間平均値を求めた後、入力音声Scのケプストラムからその時間平均値を差し引くこととしている。

【0010】

しかし、スペクトルサブトラクション法とCMN法の何れの方法も、ノイズだけを除去することは極めて困難であり、話者適応により補償すべき話者固有の発話の特徴情報も欠落してしまう場合があるため、適切な話者適応音響モデル200を作成することができなくなり、ひいては音声認識率の低下を招くという問題があった。

【0011】

本発明は上記従来の問題点を克服し、ノイズに対してロバスト (robust) で音声認識率の向上を可能にする話者適応型の音声認識システムを提供することを目的とする。

【0012】

【課題を解決するための手段】

上記目的を達成するため本発明の音声認識システムは、音声情報に基づいて生成された標準ベクトルを有する標準音響モデルと、指定テキストに対応する発話となされると、上記発話された音声からノイズを除去した第1の特徴ベクトル系列を生成する第1の特徴ベクトル生成手段と、上記発話された声からノイズを除去することなく第2の特徴ベクトル系列を生成する第2の特徴ベクトル生成手段

と、上記第1及び第2の特徴ベクトル生成手段で生成される上記第1，第2の特徴ベクトル系列に基づいて、上記標準ベクトルを補償した適応ベクトルを生成することで、上記指定テキストに対応して発話された声に適応した話者適応音響モデルを作成する作成手段とを具備することを特徴とする。

## 【0013】

また、上記作成手段は、上記第1の特徴ベクトル生成手段で生成される上記第1の特徴ベクトル系列と上記標準ベクトルとの対応関係を調べることでパス探索を行うと共に、上記第2の特徴ベクトル生成手段で生成される上記第2の特徴ベクトル系列と上記標準ベクトルとの対応付けを上記パス探索の結果に応じて行うことにより、上記適応ベクトルを生成することを特徴とする。

## 【0014】

かかる構成によれば、話者適応の際に、第1の特徴ベクトル生成手段は、話者を取りまいている周囲環境の加法性ノイズや本音声認識システム自体の伝送ノイズ等の乗法性ノイズを除いた第1の特徴ベクトルを生成し、第2の特徴ベクトル生成手段は、話者を取りまいている周囲環境の加法性ノイズや本音声認識システム自体の伝送ノイズ等の乗法性ノイズの特徴を含んだ第2の特徴ベクトルを生成する。そして、作成手段が、これらノイズを含まない第1の特徴ベクトルとノイズを含んだ第2の特徴ベクトルに基づいて標準ベクトルを補償することで適応ベクトルを生成し、この適応ベクトルによって話者の声に適応した話者適応音響モデルを生成する。

## 【0015】

このようにノイズを含まない特徴ベクトルとノイズの特徴を含んだ特徴ベクトルに基づいて、標準音響モデル中の標準ベクトルを補償することで、実際の発話環境に対応した話者適応音響モデルを作成することが可能となり、ノイズに対してロバストで音声認識率の高い音声認識システムを実現することが可能となる。

## 【0016】

また、第2の特徴ベクトル生成手段は、加法性ノイズや乗法性ノイズを除去しないで特徴ベクトルを生成し、この特徴ベクトルを話者適応に用いるので、本来の音声の特徴情報を除去することなく、適切な話者適応音響モデルを生成するこ



とができる。

【0017】

【発明の実施の形態】

以下、本発明の実施の形態を図面を参照して説明する。尚、図1は本実施形態の音声認識システムの構成を示すブロック図である。

【0018】

図1において、本音声認識システムは、隠れマルコフモデル (Hidden Markov Model: HMM) を用いて予め作成された不特定話者の標準音響モデル (以下、「標準音声HMM」という) 300と、話者適用によって生成される話者適応音響モデル (以下、「適応音声HMM」という) 400を備えて構成されている。

【0019】

尚、本実施形態を分かり易く例示するために、標準音声HMM300の状態数を1とする。また、標準音声HMM300は音節毎に出力確立分布を有しており、この出現確率分布の平均ベクトルが標準ベクトルとなっている。

【0020】

よって、標準音声HMM300は、図2に模式的に示すように、音節毎にM次元の標準ベクトル  $[a_{n,M}]$  をもっている。つまり、標準音声HMM300を作成する際、例えば静寂環境の下で収録した1人或いは複数人の話者 (不特定話者) が発話した音声データを所定時間毎にフレーム化し、そのフレーム化した音声を順次にケプストラム演算することで、音節毎に複数フレーム分のケプストラム領域での特徴ベクトル系列を生成し、更に音節毎に複数フレーム分の特徴ベクトル系列の平均を求める等によって、音節毎の標準ベクトル  $[a_{n,M}]$  からなる標準音声HMM300を作成している。

【0021】

ここで、標準ベクトル  $[a_{n,M}]$  の変数  $n$  は、各音節を識別するための状態番号、変数  $M$  はベクトルの次元を表している。例えば状態番号  $n=1$  に該当する日本語の音節 [あ] は、M次元の標準ベクトル  $[a_{n,M}] = [a_{1,1} \ a_{1,2} \ a_{1,3} \ \dots \ a_{1,M}]$ 、状態番号  $n=2$  に該当する日本語の音節 [い] は、M次元の標準ベクトル  $[a_{n,M}] = [a_{2,1} \ a_{2,2} \ a_{2,3} \ \dots \ a_{2,M}]$ 、以下同様にして、残

りの音節も状態番号  $n$  によって区別された  $M$  次元の標準ベクトル  $[a_{n,M}]$  として特徴付けられている。

#### 【 0 0 2 2 】

後述する話者適応の際に、標準音声 HMM 3 0 0 は、予め決められた文章や音節等の指定テキスト  $T_x$  が供給され、その指定テキスト  $T_x$  を構成している音節に該当する標準ベクトル  $[a_{n,M}]$  を音節の配列順に従って、パス探索部 1 0 と話者適応部 1 1 に供給する。

#### 【 0 0 2 3 】

例えば、日本語の「こんにちわ」の指定テキスト  $T_x$  が供給されると、〔こ〕、〔ん〕、〔に〕、〔ち〕、〔わ〕を表す夫々の状態番号  $n = 10, 46, 22, 17, 44$  に該当する標準ベクトル  $[a_{10,1} \ a_{10,2} \ a_{10,3} \ \dots \ a_{10,M}]$ 、 $[a_{46,1} \ a_{46,2} \ a_{46,3} \ \dots \ a_{46,M}]$ 、 $[a_{22,1} \ a_{22,2} \ a_{22,3} \ \dots \ a_{22,M}]$ 、 $[a_{17,1} \ a_{17,2} \ a_{17,3} \ \dots \ a_{17,M}]$ 、 $[a_{44,1} \ a_{44,2} \ a_{44,3} \ \dots \ a_{44,M}]$  を順番にパス探索部 1 0 と話者適応部 1 1 に供給する。

#### 【 0 0 2 4 】

更に、本音声認識システムには、フレーム化部 6 と、加法性ノイズ除去部 7、特徴ベクトル生成部 8、乗法性ノイズ除去部 9、特徴ベクトル生成部 1 2 が備えられている。

#### 【 0 0 2 5 】

フレーム化部 6 は、話者適応の際に特定話者が指定テキスト  $T_x$  を実際に発話すると、その入力音声  $S_c$  を所定時間（例えば、10～20 msec）毎のフレームに区分けして加法性ノイズ除去部 7、1 3 及び特徴ベクトル生成部 1 2 側へ出力する。

#### 【 0 0 2 6 】

加法性ノイズ除去部 7 は、フレーム毎に区分けされた各入力音声  $S_{cf}$  を順次にフーリエ変換することで、フレーム毎のスペクトルを生成し、更に各スペクトルに含まれている加法性ノイズをスペクトル領域において除去して出力する。

#### 【 0 0 2 7 】

特徴ベクトル生成部 8 は、加法性ノイズの除去されたスペクトルをフレーム毎

にケプストラム演算することで、ケプストラム領域での特徴ベクトル系列  $[c_{i,M}]'$  を生成する。尚、特徴ベクトル系列  $[c_{i,M}]'$  の変数  $i$  はフレーム毎の順番（番号）、変数  $M$  は次元を表している。

## 【0028】

乗法性ノイズ除去部 9 は、CMN 法を用いて特徴ベクトル系列  $[c_{i,M}]'$  から乗法性ノイズを除去する。すなわち、特徴ベクトル生成部 8 によってフレーム  $i$  毎に求められた複数個の特徴ベクトル系列  $[c_{i,M}]'$  を各次元毎に時間平均し、それによって求まる  $M$  次元の時間平均値  $[c_M]$  を各特徴ベクトル系列  $[c_{i,M}]'$  から減算することで、乗法性ノイズを除去した特徴ベクトル系列  $[c_{i,M}]$  を生成し、パス探索部 10 に供給する。

## 【0029】

特徴ベクトル生成部 12 は、フレーム化部 6 から出力されるフレーム毎に区分けされた各入力音声  $S_{cf}$  を順次にフーリエ変換することでフレーム毎のスペクトルを生成し、更に各スペクトルをフレーム毎にケプストラム演算することで、ケプストラム領域での特徴ベクトル系列  $[s_{i,M}]$  を生成し、話者適応部 11 に供給する。尚、特徴ベクトル系列  $[s_{i,M}]$  の変数  $i$  はフレーム毎の順番、変数  $M$  は次元を表している。

## 【0030】

こうしてパス探索部 10 には、指定テキスト  $T_x$  と標準ベクトル  $[a_{n,M}]$  と特徴ベクトル系列  $[c_{i,M}]$  が供給され、話者適応部 11 には、指定テキスト  $T_x$  と標準ベクトル  $[a_{n,M}]$  と特徴ベクトル系列  $[s_{i,M}]$  が供給される。

## 【0031】

パス探索部 10 は、標準ベクトル  $[a_{n,M}]$  と特徴ベクトル系列  $[c_{i,M}]$  とを対比し、フレーム毎の特徴ベクトル系列  $[c_{i,M}]$  が指定テキスト  $T_x$  のどの音節に該当するか判定することでパス探索を行い、そのパス探索結果  $D_v$  を話者適応部 11 に供給する。

## 【0032】

話者適応部 11 は、特徴ベクトル生成部 12 からの特徴ベクトル系列  $[s_{i,M}]$  をパス探索結果  $D_v$  に基づいて音節毎に区分けする。そして、区分けした音節

毎の特徴ベクトル系列  $[s_{i,M}]$  について次元毎に平均を求めることにより、音節毎の平均特徴ベクトル  $[s'_{n,M}]$  を生成する。

## 【0033】

更に話者適応部 11 は、指定テキスト Tx に該当する各音節の標準ベクトル  $[a_{n,M}]$  と平均特徴ベクトル  $[s'_{n,M}]$  との夫々の差ベクトル  $[d_{n,M}]$  を求める。そして、それらの差ベクトル  $[d_{n,M}]$  を平均演算することにより、特定話者の特徴を表す M 次元の移動ベクトル  $[m_M]$  を求め、更に標準音声 HMM 300 からの全ての音節の標準ベクトル  $[a_{n,M}]$  に移動ベクトル  $[m_M]$  を加えることで、全ての音節についての適応ベクトル  $[x_{n,M}]$  を生成し、それらの適応ベクトル  $[x_{n,M}]$  によって適応音声 HMM 400 を更新する。

## 【0034】

次に、パス探索部 10 と話者適応部 11 の機能を図 2～図 8 に基づいてより詳細に説明する。

## 【0035】

尚、典型例として、日本語の「こんにちわ」の指定テキスト Tx に基づいて話者適応する場合を説明する。

## 【0036】

また、話者が「こんにちわ」と発話した入力音声 Sc をフレーム化部 6 が 30 フレームに区分けして入力するものとする。

## 【0037】

標準音声 HMM 300 は、図 2 に示した様に、複数個の各音節に対応付けられた不特定話者の標準ベクトル  $[a_{n,M}]$  として作成されている。また、各音節は状態番号 n によって分類されている。

## 【0038】

また、適応音声 HMM 400 は、話者適応がなされる前は、図 2 に示した標準音声 HMM 300 の標準ベクトル  $[a_{n,M}]$  と同じ内容に設定（デフォルト設定）されている。

## 【0039】

話者適応の処理が開始され、「こんにちわ」の指定テキスト Tx が標準音声 H

MM300に供給されると、音節〔こ〕を表す状態番号 $n=10$ に該当する標準ベクトル $[a_{10,1} \ a_{10,2} \ a_{10,3} \ \cdots \ a_{10,M}]$ と、音節〔ん〕を表す状態番号 $n=46$ に該当する標準ベクトル $[a_{46,1} \ a_{46,2} \ a_{46,3} \ \cdots \ a_{46,M}]$ と、音節〔に〕を表す状態番号 $n=22$ に該当する標準ベクトル $[a_{22,1} \ a_{22,2} \ a_{22,3} \ \cdots \ a_{22,M}]$ と、音節〔ち〕を表す状態番号 $n=17$ に該当する標準ベクトル $[a_{17,1} \ a_{17,2} \ a_{17,3} \ \cdots \ a_{17,M}]$ と、音節〔わ〕を表す状態番号 $n=44$ に該当する標準ベクトル $[a_{44,1} \ a_{44,2} \ a_{44,3} \ \cdots \ a_{44,M}]$ がパス探索部10と話者適応部11に供給される。

## 【0040】

次に、特定話者が「こんにちわ」と発話すると、フレーム化部6がその入力音声 $S_c$ を時間経過に従って30フレームに区別して出力し、特徴ベクトル生成部12が、各フレームの順番に従って入力音声 $S_{cf}$ の特徴ベクトル $[s_{1,1} \ s_{1,2} \ s_{1,3} \ \cdots \ s_{1,M}] \sim [s_{30,1} \ s_{30,2} \ s_{30,3} \ \cdots \ s_{30,M}]$ を生成して話者適応部11に供給する。

## 【0041】

すなわち、図3に模式的に示すように、特徴ベクトル生成部12は、 $i=1 \sim 30$ の30フレーム分の特徴ベクトル系列 $[s_{i,M}] = [s_{1,1} \ s_{1,2} \ s_{1,3} \ \cdots \ s_{1,M}] \sim [s_{30,1} \ s_{30,2} \ s_{30,3} \ \cdots \ s_{30,M}]$ を生成して話者適応部11に供給する。

## 【0042】

一方、加法性ノイズ除去部7と特徴ベクトル生成部8と乗法性ノイズ除去部9で構成された処理系統では、フレーム化部6から供給される各フレームの入力音声 $S_{cf}$ に基づいて、 $i=1 \sim 30$ の30フレーム分の特徴ベクトル系列 $[c_{i,M}] = [c_{1,1} \ c_{1,2} \ c_{1,3} \ \cdots \ c_{1,M}] \sim [c_{30,1} \ c_{30,2} \ c_{30,3} \ \cdots \ c_{30,M}]$ を生成し、パス探索部10に供給する。すなわち、図4に模式的に示すように、30フレーム分の特徴ベクトル系列 $[c_{i,M}] = [c_{1,1} \ c_{1,2} \ c_{1,3} \ \cdots \ c_{1,M}] \sim [c_{30,1} \ c_{30,2} \ c_{30,3} \ \cdots \ c_{30,M}]$ が乗法性ノイズ除去部9を通じてパス探索部10に供給される。

## 【0043】

パス探索部10は、ビタビ (Viterbi) アルゴリズムやフォワードバックワードアルゴリズム等の手法によって、30フレーム分の特徴ベクトル系列 $[c_{i,M}]$ と指定テキストTxの各音節に対応する標準ベクトル $[a_{n,M}]$ とを対比し、特徴ベクトル系列 $[c_{i,M}]$ がフレーム毎にどの時点でどの音節に該当するか調べる。

## 【0044】

これにより、図5に示すように、30フレームの各フレーム番号*i*と、「こんにちわ」の各音節を表す各状態番号*n*との対応付けを行う。そして、この対応付けした結果をパス探索結果Dvとして話者適応部11に供給する。

## 【0045】

話者適応部11は、パス探索結果Dvに基づいて、特徴ベクトル $[s_{1,1} \ s_{1,2} \ s_{1,3} \ \dots \ s_{1,M}] \sim [s_{30,1} \ s_{30,2} \ s_{30,3} \ \dots \ s_{30,M}]$ と、標準ベクトル $[a_{10,1} \ a_{10,2} \ a_{10,3} \ \dots \ a_{10,M}]$ ,  $[a_{46,1} \ a_{46,2} \ a_{46,3} \ \dots \ a_{46,M}]$ ,  $[a_{22,1} \ a_{22,2} \ a_{22,3} \ \dots \ a_{22,M}]$ ,  $[a_{17,1} \ a_{17,2} \ a_{17,3} \ \dots \ a_{17,M}]$ ,  $[a_{44,1} \ a_{44,2} \ a_{44,3} \ \dots \ a_{44,M}]$ との対応付けを行う。

## 【0046】

すなわち、図6に示すように、パス探索によって求めた音節「こ」に該当するフレーム番号*i* = 1 ~ 6の特徴ベクトル $[s_{1,1} \ s_{1,2} \ s_{1,3} \ \dots \ s_{1,M}] \sim [s_{6,1} \ s_{6,2} \ s_{6,3} \ \dots \ s_{6,M}]$ に対して、標準ベクトル $[a_{10,1} \ a_{10,2} \ a_{10,3} \ \dots \ a_{10,M}]$ の対応付けが行われ、音節「ん」に該当するフレーム番号*i* = 7 ~ 10の特徴ベクトル $[s_{7,1} \ s_{7,2} \ s_{7,3} \ \dots \ s_{7,M}] \sim [s_{10,1} \ s_{10,2} \ s_{10,3} \ \dots \ s_{10,M}]$ に対して、標準ベクトル $[a_{46,1} \ a_{46,2} \ a_{46,3} \ \dots \ a_{46,M}]$ の対応付けが行われる。

## 【0047】

更に、音節「に」に該当するフレーム番号*i* = 11 ~ 14の特徴ベクトル $[s_{11,1} \ s_{11,2} \ s_{11,3} \ \dots \ s_{11,M}] \sim [s_{14,1} \ s_{14,2} \ s_{14,3} \ \dots \ s_{14,M}]$ に対して、標準ベクトル $[a_{22,1} \ a_{22,2} \ a_{22,3} \ \dots \ a_{22,M}]$ の対応付けが行われ、音節「ち」に該当するフレーム番号*i* = 15 ~ 18フレームの特徴ベク

トル  $[s_{15,1} \ s_{15,2} \ s_{15,3} \ \dots \ s_{15,M}] \sim [s_{18,1} \ s_{18,2} \ s_{18,3} \ \dots \ s_{18,M}]$  に対して、標準ベクトル  $[a_{17,1} \ a_{17,2} \ a_{17,3} \ \dots \ a_{17,M}]$  の対応付けが行われ、音節〔わ〕に該当するフレーム番号  $i = 19 \sim 30$  の特徴ベクトル  $[s_{19,1} \ s_{19,2} \ s_{19,3} \ \dots \ s_{19,M}] \sim [s_{30,1} \ s_{30,2} \ s_{30,3} \ \dots \ s_{30,M}]$  に対して、標準ベクトル  $[a_{44,1} \ a_{44,2} \ a_{44,3} \ \dots \ a_{44,M}]$  の対応付けが行われる。

【0048】

次に、話者適応部 11 は、図 6 に示した 30 フレーム分の特徴ベクトル  $[s_{1,1} \ s_{1,2} \ s_{1,3} \ \dots \ s_{1,M}] \sim [s_{30,1} \ s_{30,2} \ s_{30,3} \ \dots \ s_{30,M}]$  を各音節〔こ〕，〔ん〕，〔に〕，〔ち〕，〔わ〕毎に区分けし、区分けした特徴ベクトル毎の平均を求めることにより、図 7 に示すように、各音節〔こ〕，〔ん〕，〔に〕，〔ち〕，〔わ〕毎の平均特徴ベクトル  $[s'_{n,M}]$  を生成する。

【0049】

つまり、図 6 に示した音節〔こ〕に該当する第 1 ～第 6 フレーム（フレーム数  $k = 6$ ）の特徴ベクトル  $[s_{1,1} \ s_{1,2} \ s_{1,3} \ \dots \ s_{1,M}] \sim [s_{6,1} \ s_{6,2} \ s_{6,3} \ \dots \ s_{6,M}]$  については、次式（1）で示すように、第 1 次元目の 6 個の要素  $s_{1,1} \sim s_{6,1}$  を加算し、その加算値  $(s_{1,1} + s_{2,1} + s_{3,1} + s_{4,1} + s_{5,1} + s_{6,1})$  をフレーム数  $k = 6$  で乗算することにより、平均特徴ベクトル  $[s'_{n,M}]$  の第 1 次元目の要素  $s'_{n,1}$  を求める。また、第 2 次元目の 6 個の要素  $s_{1,2} \sim s_{6,2}$  についても同様に、加算値  $(s_{1,2} + s_{2,2} + s_{3,2} + s_{4,2} + s_{5,2} + s_{6,2})$  を求め、フレーム数  $k = 6$  で乗算することにより、平均特徴ベクトル  $[s'_{n,M}]$  の第 2 次元目の要素  $s'_{n,2}$  を求める。以下同様にして、第 M 次元目の 6 個の要素  $s_{1,M} \sim s_{6,M}$  までの要素  $s'_{n,M}$  を求め、これら M 次元の要素  $s'_{n,1} \sim s'_{n,M}$  からなる音節〔こ〕に該当する M 次元の平均特徴ベクトル  $[s'_{n,1} \ s'_{n,2} \ s'_{n,3} \ \dots \ s'_{n,M}]$  を生成する。

【0050】

【数 1】

$$s'_{n,M} = (s_{1,1} + s_{2,1} + s_{3,1} + s_{4,1} + s_{5,1} + s_{6,1}) / k \quad \dots \quad (1)$$

尚、式(1)中の変数 $k$ は各音節におけるフレーム数、変数 $n$ は各音節を識別するための状態番号、変数 $M$ は次元を表している。したがって、式(1)中の変数 $n$ は、 $n=10$ となり、音節「こ」に該当する $M$ 次元の平均特徴ベクトルは $[s_{10,1}^{\wedge} \ s_{10,2}^{\wedge} \ s_{10,3}^{\wedge} \ \cdots \ s_{10,M}^{\wedge}]$ となる。

また、残りの音節「ん」に該当する平均特徴ベクトル $[s_{46,1}^{\wedge} \ \cdots \ s_{46,M}^{\wedge}]$ と、音節「に」に該当する平均特徴ベクトル $[s_{22,1}^{\wedge} \ \cdots \ s_{22,M}^{\wedge}]$ と、音節「ち」に該当する平均特徴ベクトル $[s_{17,1}^{\wedge} \ \cdots \ s_{17,M}^{\wedge}]$ と、音節「わ」に該当する平均特徴ベクトル $[s_{44,1}^{\wedge} \ \cdots \ s_{44,M}^{\wedge}]$ も同様にして求める。

【0051】

次に、次式(2)に基づいて、各音節「こ」, 「ん」, 「に」, 「ち」, 「わ」に対応する平均特徴ベクトル $[s_{10,1}^{\wedge} \ \cdots \ s_{10,M}^{\wedge}]$ ,  $[s_{46,1}^{\wedge} \ \cdots \ s_{46,M}^{\wedge}]$ ,  $[s_{22,1}^{\wedge} \ \cdots \ s_{22,M}^{\wedge}]$ ,  $[s_{17,1}^{\wedge} \ \cdots \ s_{17,M}^{\wedge}]$ ,  $[s_{44,1}^{\wedge} \ \cdots \ s_{44,M}^{\wedge}]$ と、標準ベクトル $[a_{10,1} \ \cdots \ a_{10,M}]$ ,  $[a_{46,1} \ \cdots \ a_{46,M}]$ ,  $[a_{22,1} \ \cdots \ a_{22,M}]$ ,  $[a_{17,1} \ \cdots \ a_{17,M}]$ ,  $[a_{44,1} \ \cdots \ a_{44,M}]$ とのそれぞれ差ベクトル $[d_{10,1} \ \cdots \ d_{10,M}]$ ,  $[d_{46,1} \ \cdots \ d_{46,M}]$ ,  $[d_{22,1} \ \cdots \ d_{22,M}]$ ,  $[d_{17,1} \ \cdots \ d_{17,M}]$ ,  $[d_{44,1} \ \cdots \ d_{44,M}]$ を求める。

【0052】

【数2】

$$d_{n,j} = s_{n,j}^{\wedge} - a_{n,j} \quad \cdots \quad (2)$$

尚、上記式(2)中の変数 $n$ は、各音節「こ」, 「ん」, 「に」, 「ち」, 「わ」に対応する状態番号 $n=10, 46, 22, 17, 44$ を示し、変数 $j$ は、ベクトルの各次元 $j=1 \sim M$ を示している。

【0053】

そして、求めた差ベクトル $[d_{10,1} \ \cdots \ d_{10,M}]$ ,  $[d_{46,1} \ \cdots \ d_{46,M}]$ ,  $[d_{22,1} \ \cdots \ d_{22,M}]$ ,  $[d_{17,1} \ \cdots \ d_{17,M}]$ ,  $[d_{44,1} \ \cdots \ d_{44,M}]$ を次式(3)に適用して、各次元毎の平均を求めることにより、これら5個( $V=5$ )の音節「こ」, 「ん」, 「に」, 「ち」, 「わ」の $M$ 次元移動ベクトル $[m_M] = [m_1, m_2, \cdots, m_M]$ を求める。



【0054】

【数3】

$$m_j = \frac{1}{V} \sum_n d_{n,j} \quad \dots (3)$$

尚、上記(3)中の変数 $j$ は、ベクトルの各次元 $j = 1 \sim M$ を示している。また、変数 $n$ は、各音節「こ」、「ん」、「に」、「ち」、「わ」に対応する状態番号 $n = 10, 46, 22, 17, 44$ を示し、変数 $V$ は音節の数( $V = 5$ )を示している。

【0055】

こうして求めた移動ベクトル $[m_M] = [m_1, m_2, \dots, m_M]$ は、特定話者の特徴を表している。そこで、次式(4)の演算式で示すように、全ての音節の標準ベクトル $[a_{n,M}]$ に移動ベクトル $[m_M]$ を加えることで、話者固有の特徴を有する適応ベクトル $[x_{n,M}]$ を求め、更に、図8に示すように、求めた適応ベクトル $[x_{n,M}]$ によって適応音声HMM400を更新することで、話者適応の処理を完了する。

【0056】

【数4】

$$[x_{n,M}] = [a_{n,M}] + [m_M] \quad \dots (4)$$

また、一例として「こんにちわ」の指定テキスト $T_x$ に基づいて適応音声HMM400を話者適応する場合を説明したが、他の音節を含んだ指定テキスト $T_x$ に基づいて適応音声HMM400を話者適応することで、適応音声HMM400中の全ての音節についても話者適応が行われるようになっている。

【0057】

次に、話者適応により適応音声HMM400を生成した後、特定話者が任意の発話をする、その入力音声 $S_c$ をフレーム化部6が上記同様に所定時間(例えば、10～20msec)毎のフレームに区分けし、各フレームの入力音声 $S_{cf}$ を時間経過順に出力して加法性ノイズ除去部13側へ供給する。

【0058】

加法性ノイズ除去部 13 は、上記した加法性ノイズ除去部 7 と同様に、フレーム毎に区分けされた各入力音声  $S_{cf}$  を順次にフーリエ変換することでフレーム毎のスペクトルを生成し、更に各スペクトルに含まれている加法性ノイズをスペクトル領域において除去して、特徴ベクトル生成部 14 側へ出力する。

【0059】

特徴ベクトル生成部 14 は、上記した特徴ベクトル生成部 8 と同様に、加法性ノイズの除去されたスペクトルをフレーム毎にケプストラム演算することで、ケプストラム領域での特徴ベクトル系列  $[y_{i,M}]'$  を生成して、乗法性ノイズ除去部 15 側へ出力する。

【0060】

乗法性ノイズ除去部 15 は、上記した乗法性ノイズ除去部 9 と同様に、CMN 法を用いて特徴ベクトル系列  $[y_{i,M}]'$  から乗法性ノイズを除去し、その乗法性ノイズを除去した M 次元の特徴ベクトル系列  $[y_{i,M}]$  を認識部 16 に供給する。尚、特徴ベクトル系列  $[y_{i,M}]$  の変数  $i$  はフレーム番号を表している。

【0061】

こうして、実際に発話された声の特徴ベクトル系列  $[y_{i,M}]$  が認識部 16 に供給されると、認識部 16 は、特徴ベクトル系列  $[y_{i,M}]$  と話者適応した適応音声 HMM400 の適応ベクトル  $[x_{n,M}]$  とを照合し、最大尤度を与える適応音声 HMM400 を認識結果として出力する。

【0062】

以上に述べたように本実施形態の音声認識システムによれば、話者適応の際に、指定テキスト  $T_x$  に従って特定話者が発話すると、加法性ノイズ除去部 7 と特徴ベクトル生成部 8 及び乗法性ノイズ除去部 9 が、加法性ノイズと乗法性ノイズを除去した特徴ベクトル系列  $[c_{i,M}]$  を生成すると共に、特徴ベクトル生成部 12 が、加法性ノイズと乗法性ノイズを含んでいる入力音声  $S_{cf}$  に基づいて特徴ベクトル系列  $[s_{i,M}]$  を生成し、パス探索部 10 と話者適応部 11 が、これらの特徴ベクトル系列  $[c_{i,M}]$  と特徴ベクトル系列  $[s_{i,M}]$  及び標準ベクトル  $[a_{i,M}]$  に基づいて話者適応した適応ベクトル  $[x_{i,M}]$  を生成し、その話者適応した適応ベクトル  $[x_{i,M}]$  で適応音声 HMM400 を更新している。

## 【0063】

このため、特定話者を取りまいてる周囲環境のノイズ（加法性ノイズ）や本音声認識システム自体の伝送ノイズ（乗法性ノイズ）の特徴を含んだ特徴ベクトル系列  $[s_{i,M}]$  が話者適応に用いられることになり、実際の発話環境に対応した適応音声HMM400を生成することが可能となることから、ノイズに対してロバストで音声認識率の高い音声認識システムを実現することができる。

## 【0064】

また、従来の話者適応型の音声認識システムでは、話者適応の際に、加法性ノイズや乗法性ノイズを除去した特徴ベクトルを生成するのに伴って、話者適応により補償すべき話者固有の発話の特徴情報も欠落してしまい、適切な話者適応音響モデルを作成することができないという問題があったが、本実施形態の音声認識システムによれば、特徴ベクトル生成部12が、加法性ノイズや乗法性ノイズを除去しないで特徴ベクトル系列  $[s_{i,M}]$  を生成し、この特徴ベクトル系列  $[s_{i,M}]$  を話者適応に用いるので、話者適応により補償すべき話者固有の発話の特徴情報を欠落することなく、適切な話者適応音響モデルを作成することができ、音声認識率の向上を図ることができる。

## 【0065】

尚、本実施形態では、日本語の「あいうえお」等の音節を基準にした適応音声HMM400を作成する場合を説明したが、音節のみに限らず、音素等を基準として適応音声HMM400を作成することが可能である。

## 【0066】

また、本実施形態では、簡単な例示を上げて、話者適応の手法について説明したが、標準ベクトル  $[a_{n,M}]$  と適応話者の特徴ベクトル系列  $[s_{i,M}]$  或いは  $[c_{i,M}]$  との対応を求め、それに基づいて話者適応音響モデルを生成する他の様々な話者適応手法に本発明の話者適応手法を適応することが可能である。

## 【0067】

## 【発明の効果】

以上説明したように本発明の音声認識システムによれば、話者適応の際に、加法性ノイズや乗法性ノイズを除いた特徴ベクトルと、加法性ノイズや乗法性ノイ

ズの特徴を含んだ特徴ベクトルとを生成し、これらノイズを含まない特徴ベクトルとノイズを含んだ特徴ベクトルに基づいて、標準ベクトルを補償することで、話者固有の発話に適応した話者適応音響モデルを作成するようにしたので、実際の発話環境に対応した話者適応音響モデルを生成することが可能となる。

【0068】

また、加法性ノイズや乗法性ノイズを除去しないで特徴ベクトルを話者適応に用いるので、話者適応により補償すべき話者固有の発話の特徴情報を欠落することなく、適切な話者適応音響モデルを生成することができる。

【0069】

このため、ノイズに対してロバストで音声認識率の高い音声認識システムを実現することが可能となる。

【図面の簡単な説明】

【図1】

本実施形態の音声認識システムの構成を示すブロック図である。

【図2】

標準音響モデルの構成を模式的に示した図である。

【図3】

話者適応の際に特徴ベクトル生成部12で生成される特徴ベクトル系列 $[s_{i,M}]$ を示した図である。

【図4】

話者適応の際に乗法性ノイズ除去部9から出力される特徴ベクトル系列 $[c_{i,M}]$ を示した図である。

【図5】

特徴ベクトル系列 $[c_{i,M}]$ と標準ベクトル $[a_{n,M}]$ との対応関係をフレーム番号と状態番号に基づいて示した図である。

【図6】

特徴ベクトル系列 $[c_{i,M}]$ と標準ベクトル $[a_{n,M}]$ とフレーム番号及び状態番号を対応付けて示した図である。

【図7】

話者適応によって生成される平均特徴ベクトルと標準ベクトルとの対応関係を示した図である。

【図 8】

更新後の話者適応音響モデルの内容を示した図である。

【図 9】

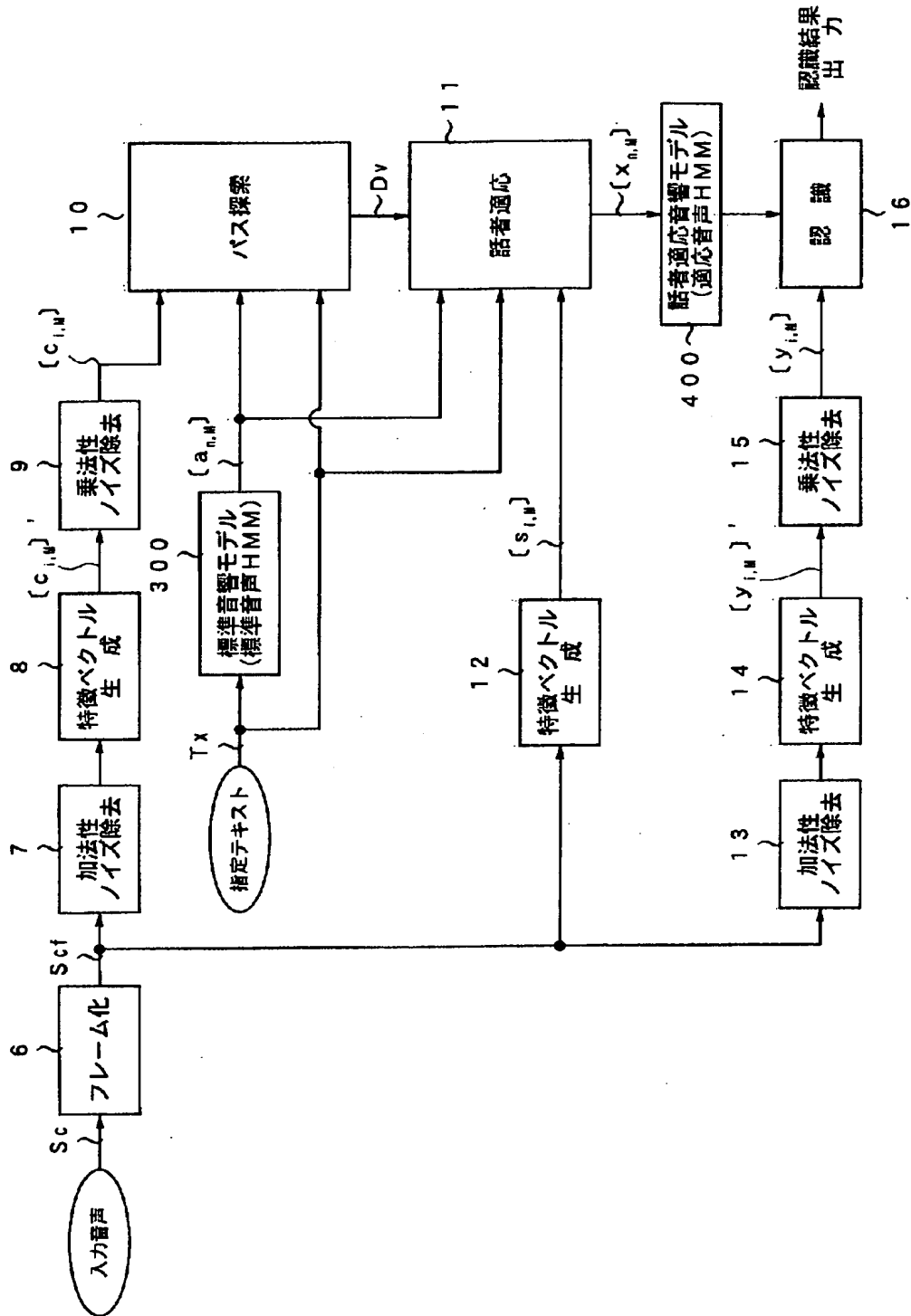
従来の話者適応型音声認識システムの構成を示すブロック図である。

【符号の説明】

- 6 … フレーム化部
- 7, 13 … 加法性ノイズ除去部
- 8, 12, 14 … 特徴ベクトル生成部
- 9, 15 … 乗法性ノイズ除去部
- 10 … パス探索部
- 11 … 話者適応部
- 16 … 認識部
- 300 … 標準音響モデル（標準音声HMM）
- 400 … 話者適応音響モデル（適応音声HMM）

【書類名】 図面

【図 1】



【図 2】

状態番号 n	音節	標準ベクトル $[a_{n, m}]$
1	あ	$[a_{1, 1} \quad a_{1, 2} \quad a_{1, 3} \quad \dots \quad a_{1, M}]$
2	い	$[a_{2, 1} \quad a_{2, 2} \quad a_{2, 3} \quad \dots \quad a_{2, M}]$
3	う	$[a_{3, 1} \quad a_{3, 2} \quad a_{3, 3} \quad \dots \quad a_{3, M}]$
4	え	$[a_{4, 1} \quad a_{4, 2} \quad a_{4, 3} \quad \dots \quad a_{4, M}]$
5	お	$[a_{5, 1} \quad a_{5, 2} \quad a_{5, 3} \quad \dots \quad a_{5, M}]$
6	か	$[a_{6, 1} \quad a_{6, 2} \quad a_{6, 3} \quad \dots \quad a_{6, M}]$
⋮	⋮	⋮
10	こ	$[a_{10, 1} \quad a_{10, 2} \quad a_{10, 3} \quad \dots \quad a_{10, M}]$
11	さ	$[a_{11, 1} \quad a_{11, 2} \quad a_{11, 3} \quad \dots \quad a_{11, M}]$
⋮	⋮	⋮
16	た	$[a_{16, 1} \quad a_{16, 2} \quad a_{16, 3} \quad \dots \quad a_{16, M}]$
17	ち	$[a_{17, 1} \quad a_{17, 2} \quad a_{17, 3} \quad \dots \quad a_{17, M}]$
⋮	⋮	⋮
21	な	$[a_{21, 1} \quad a_{21, 2} \quad a_{21, 3} \quad \dots \quad a_{21, M}]$
22	に	$[a_{22, 1} \quad a_{22, 2} \quad a_{22, 3} \quad \dots \quad a_{22, M}]$
⋮	⋮	⋮
26	は	$[a_{26, 1} \quad a_{26, 2} \quad a_{26, 3} \quad \dots \quad a_{26, M}]$
⋮	⋮	⋮
31	ま	$[a_{31, 1} \quad a_{31, 2} \quad a_{31, 3} \quad \dots \quad a_{31, M}]$
⋮	⋮	⋮
36	や	$[a_{36, 1} \quad a_{36, 2} \quad a_{36, 3} \quad \dots \quad a_{36, M}]$
37	ゆ	$[a_{37, 1} \quad a_{37, 2} \quad a_{37, 3} \quad \dots \quad a_{37, M}]$
38	よ	$[a_{38, 1} \quad a_{38, 2} \quad a_{38, 3} \quad \dots \quad a_{38, M}]$
39	ら	$[a_{39, 1} \quad a_{39, 2} \quad a_{39, 3} \quad \dots \quad a_{39, M}]$
⋮	⋮	⋮
44	わ	$[a_{44, 1} \quad a_{44, 2} \quad a_{44, 3} \quad \dots \quad a_{44, M}]$
45	を	$[a_{45, 1} \quad a_{45, 2} \quad a_{45, 3} \quad \dots \quad a_{45, M}]$
46	ん	$[a_{46, 1} \quad a_{46, 2} \quad a_{46, 3} \quad \dots \quad a_{46, M}]$
⋮	⋮	⋮
N	⋮	$[a_{N, 1} \quad a_{N, 2} \quad a_{N, 3} \quad \dots \quad a_{N, M}]$

【図 3】

時間  
↓

フレーム番号 $i$	話者適応の際に特徴ベクトル生成部 1 2 で生成される特徴ベクトル $[S_{i, M}]$
1	$[S_{1, 1} \quad S_{1, 2} \quad S_{1, 3} \quad \cdots \quad S_{1, M}]$
2	$[S_{2, 1} \quad S_{2, 2} \quad S_{2, 3} \quad \cdots \quad S_{2, M}]$
3	$[S_{3, 1} \quad S_{3, 2} \quad S_{3, 3} \quad \cdots \quad S_{3, M}]$
4	$[S_{4, 1} \quad S_{4, 2} \quad S_{4, 3} \quad \cdots \quad S_{4, M}]$
5	$[S_{5, 1} \quad S_{5, 2} \quad S_{5, 3} \quad \cdots \quad S_{5, M}]$
6	$[S_{6, 1} \quad S_{6, 2} \quad S_{6, 3} \quad \cdots \quad S_{6, M}]$
7	$[S_{7, 1} \quad S_{7, 2} \quad S_{7, 3} \quad \cdots \quad S_{7, M}]$
8	$[S_{8, 1} \quad S_{8, 2} \quad S_{8, 3} \quad \cdots \quad S_{8, M}]$
9	$[S_{9, 1} \quad S_{9, 2} \quad S_{9, 3} \quad \cdots \quad S_{9, M}]$
10	$[S_{10, 1} \quad S_{10, 2} \quad S_{10, 3} \quad \cdots \quad S_{10, M}]$
11	$[S_{11, 1} \quad S_{11, 2} \quad S_{11, 3} \quad \cdots \quad S_{11, M}]$
12	$[S_{12, 1} \quad S_{12, 2} \quad S_{12, 3} \quad \cdots \quad S_{12, M}]$
13	$[S_{13, 1} \quad S_{13, 2} \quad S_{13, 3} \quad \cdots \quad S_{13, M}]$
14	$[S_{14, 1} \quad S_{14, 2} \quad S_{14, 3} \quad \cdots \quad S_{14, M}]$
15	$[S_{15, 1} \quad S_{15, 2} \quad S_{15, 3} \quad \cdots \quad S_{15, M}]$
16	$[S_{16, 1} \quad S_{16, 2} \quad S_{16, 3} \quad \cdots \quad S_{16, M}]$
17	$[S_{17, 1} \quad S_{17, 2} \quad S_{17, 3} \quad \cdots \quad S_{17, M}]$
18	$[S_{18, 1} \quad S_{18, 2} \quad S_{18, 3} \quad \cdots \quad S_{18, M}]$
19	$[S_{19, 1} \quad S_{19, 2} \quad S_{19, 3} \quad \cdots \quad S_{19, M}]$
20	$[S_{20, 1} \quad S_{20, 2} \quad S_{20, 3} \quad \cdots \quad S_{20, M}]$
⋮	⋮
29	$[S_{29, 1} \quad S_{29, 2} \quad S_{29, 3} \quad \cdots \quad S_{29, M}]$
30	$[S_{30, 1} \quad S_{30, 2} \quad S_{30, 3} \quad \cdots \quad S_{30, M}]$



【図4】

時間 ↓	フレーム番号 i	話者適応の際に乗法性ノイズ除去部9 から出力される特徴ベクトル $[C_{i, M}]$
	1	$[C_{1, 1} \quad C_{1, 2} \quad C_{1, 3} \quad \dots \quad C_{1, M}]$
	2	$[C_{2, 1} \quad C_{2, 2} \quad C_{2, 3} \quad \dots \quad C_{2, M}]$
	3	$[C_{3, 1} \quad C_{3, 2} \quad C_{3, 3} \quad \dots \quad C_{3, M}]$
	4	$[C_{4, 1} \quad C_{4, 2} \quad C_{4, 3} \quad \dots \quad C_{4, M}]$
	5	$[C_{5, 1} \quad C_{5, 2} \quad C_{5, 3} \quad \dots \quad C_{5, M}]$
	6	$[C_{6, 1} \quad C_{6, 2} \quad C_{6, 3} \quad \dots \quad C_{6, M}]$
	7	$[C_{7, 1} \quad C_{7, 2} \quad C_{7, 3} \quad \dots \quad C_{7, M}]$
	8	$[C_{8, 1} \quad C_{8, 2} \quad C_{8, 3} \quad \dots \quad C_{8, M}]$
	9	$[C_{9, 1} \quad C_{9, 2} \quad C_{9, 3} \quad \dots \quad C_{9, M}]$
	10	$[C_{10, 1} \quad C_{10, 2} \quad C_{10, 3} \quad \dots \quad C_{10, M}]$
	11	$[C_{11, 1} \quad C_{11, 2} \quad C_{11, 3} \quad \dots \quad C_{11, M}]$
	12	$[C_{12, 1} \quad C_{12, 2} \quad C_{12, 3} \quad \dots \quad C_{12, M}]$
	13	$[C_{13, 1} \quad C_{13, 2} \quad C_{13, 3} \quad \dots \quad C_{13, M}]$
	14	$[C_{14, 1} \quad C_{14, 2} \quad C_{14, 3} \quad \dots \quad C_{14, M}]$
	15	$[C_{15, 1} \quad C_{15, 2} \quad C_{15, 3} \quad \dots \quad C_{15, M}]$
	16	$[C_{16, 1} \quad C_{16, 2} \quad C_{16, 3} \quad \dots \quad C_{16, M}]$
	17	$[C_{17, 1} \quad C_{17, 2} \quad C_{17, 3} \quad \dots \quad C_{17, M}]$
	18	$[C_{18, 1} \quad C_{18, 2} \quad C_{18, 3} \quad \dots \quad C_{18, M}]$
	19	$[C_{19, 1} \quad C_{19, 2} \quad C_{19, 3} \quad \dots \quad C_{19, M}]$
	20	$[C_{20, 1} \quad C_{20, 2} \quad C_{20, 3} \quad \dots \quad C_{20, M}]$
	⋮	⋮
	29	$[C_{29, 1} \quad C_{29, 2} \quad C_{29, 3} \quad \dots \quad C_{29, M}]$
	30	$[C_{30, 1} \quad C_{30, 2} \quad C_{30, 3} \quad \dots \quad C_{30, M}]$

【図 5】

時間  
↓

フレーム番号 i	状態番号 n	
1	1 0	{ [こ]
2	1 0	
3	1 0	
4	1 0	
5	1 0	
6	1 0	{ [ん]
7	4 6	
8	4 6	
9	4 6	
1 0	4 6	{ [に]
1 1	2 2	
1 2	2 2	
1 3	2 2	
1 4	2 2	{ [ち]
1 5	1 7	
1 6	1 7	
1 7	1 7	
1 8	1 7	{ [わ]
1 9	4 4	
2 0	4 4	
⋮	⋮	
2 9	4 4	
3 0	4 4	

【図6】

フレーム番号 $i$	特徴ベクトル $[s_{i,1} \dots s_{i,M}]$	状態番号 $n$	標準ベクトル $[a_{n,1} \dots a_{n,M}]$	
1	$[s_{1,1} \dots s_{1,M}]$	10	$[a_{10,1} \dots a_{10,M}]$	{こ}
2	$[s_{2,1} \dots s_{2,M}]$	10	$[a_{10,1} \dots a_{10,M}]$	
3	$[s_{3,1} \dots s_{3,M}]$	10	$[a_{10,1} \dots a_{10,M}]$	
4	$[s_{4,1} \dots s_{4,M}]$	10	$[a_{10,1} \dots a_{10,M}]$	
5	$[s_{5,1} \dots s_{5,M}]$	10	$[a_{10,1} \dots a_{10,M}]$	
6	$[s_{6,1} \dots s_{6,M}]$	10	$[a_{10,1} \dots a_{10,M}]$	
7	$[s_{7,1} \dots s_{7,M}]$	46	$[a_{46,1} \dots a_{46,M}]$	{ん}
8	$[s_{8,1} \dots s_{8,M}]$	46	$[a_{46,1} \dots a_{46,M}]$	
9	$[s_{9,1} \dots s_{9,M}]$	46	$[a_{46,1} \dots a_{46,M}]$	
10	$[s_{10,1} \dots s_{10,M}]$	46	$[a_{46,1} \dots a_{46,M}]$	
11	$[s_{11,1} \dots s_{11,M}]$	22	$[a_{22,1} \dots a_{22,M}]$	{に}
12	$[s_{12,1} \dots s_{12,M}]$	22	$[a_{22,1} \dots a_{22,M}]$	
13	$[s_{13,1} \dots s_{13,M}]$	22	$[a_{22,1} \dots a_{22,M}]$	
14	$[s_{14,1} \dots s_{14,M}]$	22	$[a_{22,1} \dots a_{22,M}]$	
15	$[s_{15,1} \dots s_{15,M}]$	17	$[a_{17,1} \dots a_{17,M}]$	{ち}
16	$[s_{16,1} \dots s_{16,M}]$	17	$[a_{17,1} \dots a_{17,M}]$	
17	$[s_{17,1} \dots s_{17,M}]$	17	$[a_{17,1} \dots a_{17,M}]$	
18	$[s_{18,1} \dots s_{18,M}]$	17	$[a_{17,1} \dots a_{17,M}]$	
19	$[s_{19,1} \dots s_{19,M}]$	44	$[a_{44,1} \dots a_{44,M}]$	{わ}
20	$[s_{20,1} \dots s_{20,M}]$	44	$[a_{44,1} \dots a_{44,M}]$	
⋮	⋮	⋮	⋮	
29	$[s_{29,1} \dots s_{29,M}]$	44	$[a_{44,1} \dots a_{44,M}]$	
30	$[s_{30,1} \dots s_{30,M}]$	44	$[a_{44,1} \dots a_{44,M}]$	

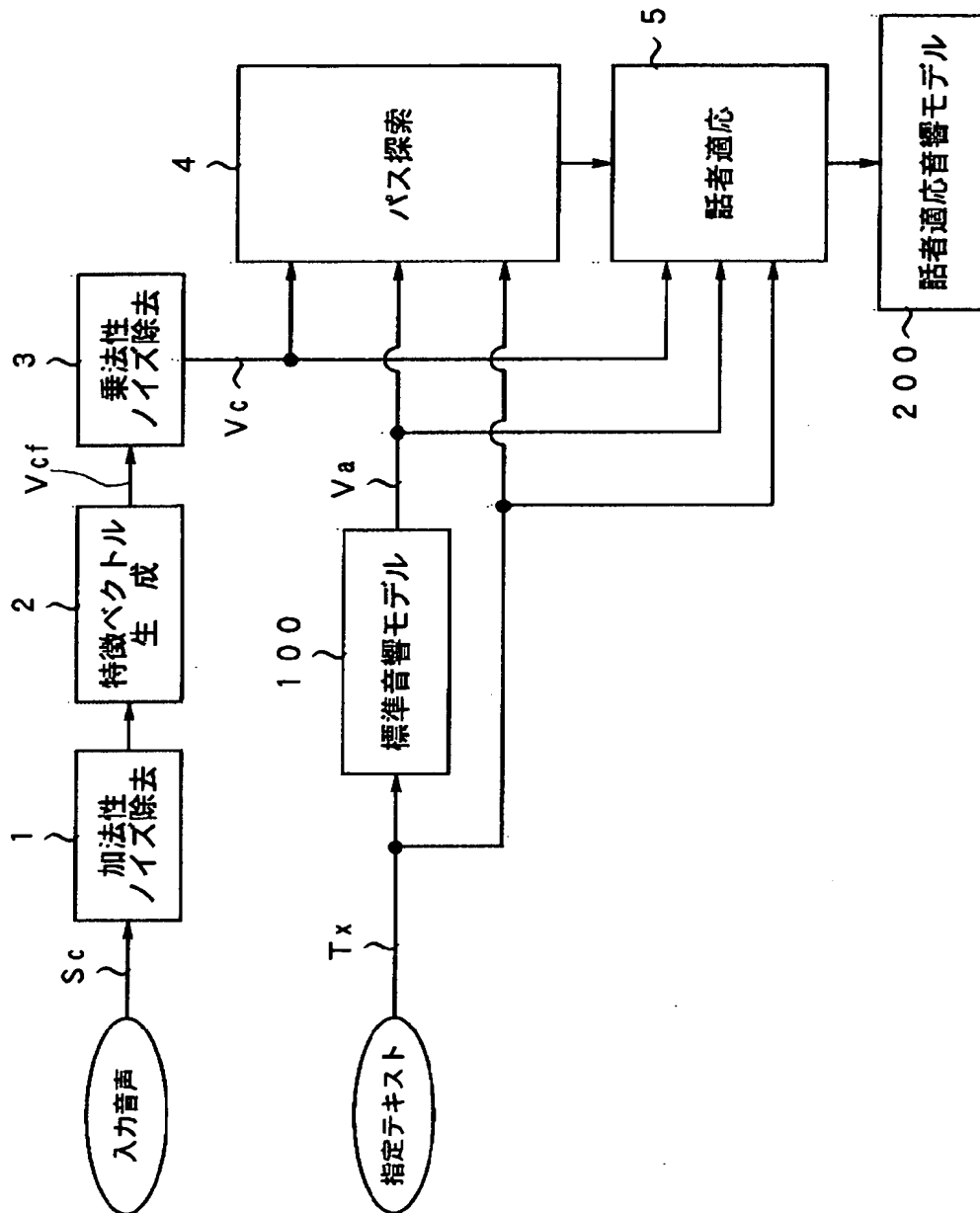
【図 7】

	平均特徴ベクトル [ $\hat{s}_{n, m}$ ]	状態番号 n	標準ベクトル [ $a_{n, m}$ ]	
i = 1 ~ 6 の平均	[ $\hat{s}_{10, 1} \cdots \hat{s}_{10, M}$ ]	1 0	[ $a_{10, 1} \cdots a_{10, M}$ ]	← (こ)
i = 7 ~ 10 の平均	[ $\hat{s}_{46, 1} \cdots \hat{s}_{46, M}$ ]	4 6	[ $a_{46, 1} \cdots a_{46, M}$ ]	← (ん)
i = 11 ~ 14 の平均	[ $\hat{s}_{22, 1} \cdots \hat{s}_{22, M}$ ]	2 2	[ $a_{22, 1} \cdots a_{22, M}$ ]	← (に)
i = 15 ~ 18 の平均	[ $\hat{s}_{17, 1} \cdots \hat{s}_{17, M}$ ]	1 7	[ $a_{17, 1} \cdots a_{17, M}$ ]	← (ち)
i = 19 ~ 30 の平均	[ $\hat{s}_{44, 1} \cdots \hat{s}_{44, M}$ ]	4 4	[ $a_{44, 1} \cdots a_{44, M}$ ]	← (わ)

【図 8】

状態番号 n	音節	更新後の更新音声HMMの適応ベクトル
1	あ	[X <sub>1,1</sub> X <sub>1,2</sub> X <sub>1,3</sub> ..... X <sub>1,M</sub> ]
2	い	[X <sub>2,1</sub> X <sub>2,2</sub> X <sub>2,3</sub> ..... X <sub>2,M</sub> ]
3	う	[X <sub>3,1</sub> X <sub>3,2</sub> X <sub>3,3</sub> ..... X <sub>3,M</sub> ]
4	え	[X <sub>4,1</sub> X <sub>4,2</sub> X <sub>4,3</sub> ..... X <sub>4,M</sub> ]
5	お	[X <sub>5,1</sub> X <sub>5,2</sub> X <sub>5,3</sub> ..... X <sub>5,M</sub> ]
6	か	[X <sub>6,1</sub> X <sub>6,2</sub> X <sub>6,3</sub> ..... X <sub>6,M</sub> ]
⋮	⋮	⋮
10	こ	[X <sub>10,1</sub> X <sub>10,2</sub> X <sub>10,3</sub> ..... X <sub>10,M</sub> ]
11	さ	[X <sub>11,1</sub> X <sub>11,2</sub> X <sub>11,3</sub> ..... X <sub>11,M</sub> ]
⋮	⋮	⋮
16	た	[X <sub>16,1</sub> X <sub>16,2</sub> X <sub>16,3</sub> ..... X <sub>16,M</sub> ]
17	ち	[X <sub>17,1</sub> X <sub>17,2</sub> X <sub>17,3</sub> ..... X <sub>17,M</sub> ]
⋮	⋮	⋮
21	な	[X <sub>21,1</sub> X <sub>21,2</sub> X <sub>21,3</sub> ..... X <sub>21,M</sub> ]
22	に	[X <sub>22,1</sub> X <sub>22,2</sub> X <sub>22,3</sub> ..... X <sub>22,M</sub> ]
⋮	⋮	⋮
26	は	[X <sub>26,1</sub> X <sub>26,2</sub> X <sub>26,3</sub> ..... X <sub>26,M</sub> ]
⋮	⋮	⋮
31	ま	[X <sub>31,1</sub> X <sub>31,2</sub> X <sub>31,3</sub> ..... X <sub>31,M</sub> ]
⋮	⋮	⋮
36	や	[X <sub>36,1</sub> X <sub>36,2</sub> X <sub>36,3</sub> ..... X <sub>36,M</sub> ]
37	ゆ	[X <sub>37,1</sub> X <sub>37,2</sub> X <sub>37,3</sub> ..... X <sub>37,M</sub> ]
38	よ	[X <sub>38,1</sub> X <sub>38,2</sub> X <sub>38,3</sub> ..... X <sub>38,M</sub> ]
39	ら	[X <sub>39,1</sub> X <sub>39,2</sub> X <sub>39,3</sub> ..... X <sub>39,M</sub> ]
⋮	⋮	⋮
44	わ	[X <sub>44,1</sub> X <sub>44,2</sub> X <sub>44,3</sub> ..... X <sub>44,M</sub> ]
45	を	[X <sub>45,1</sub> X <sub>45,2</sub> X <sub>45,3</sub> ..... X <sub>45,M</sub> ]
46	ん	[X <sub>46,1</sub> X <sub>46,2</sub> X <sub>46,3</sub> ..... X <sub>46,M</sub> ]
⋮	⋮	⋮
N	:	[X <sub>N,1</sub> X <sub>N,2</sub> X <sub>N,3</sub> ..... X <sub>N,M</sub> ]

【図 9】



【書類名】 要約書

【要約】

【課題】 ノイズに対してロバストな音声認識システムを提供する。

【解決手段】 話者適応の際に、第1の特徴ベクトル生成手段(7, 8, 9)によって、加法性ノイズと乗法性ノイズを除いた特徴ベクトル系列 $[c_{i,M}]$ を生成すると共に、第2の特徴ベクトル生成手段(12)によって、加法性ノイズと乗法性ノイズの特徴を含んだ特徴ベクトル系列 $[s_{i,M}]$ を生成する。パス探索部(10)が、特徴ベクトル系列 $[c_{i,M}]$ と標準音声HMM(300)の標準ベクトル $[a_{n,M}]$ とを対比することでパス探索を行い、話者適応部(11)が、そのパス探索結果 $D_v$ に対応する標準ベクトル $[a_{n,M}]$ と特徴ベクトル系列 $[s_{i,M}]$ の平均特徴ベクトル $[s'_{n,M}]$ とを相関演算することにより、適応ベクトル $[x_{n,M}]$ を生成し、音声認識に用いる話者適応音響モデル(400)の特徴ベクトルを適応ベクトル $[x_{n,M}]$ で更新する。

【選択図】 図1

出 願 人 履 歴 情 報

識別番号 [000005016]

1. 変更年月日	1990年 8月31日
[変更理由]	新規登録
住 所	東京都目黒区目黒1丁目4番1号
氏 名	パイオニア株式会社